SEEDS

Workshop Paper #2:

Data Service Provider Types - A Working List

January 16, 2002

G. Hunolt, SGT, Inc.

Question for the Workshop:

1. Are the seven ESE data service provider types representative of the different types of institutional components / data centers / projects that will be implemented in the SEEDS era? If not, can you suggest others?

Outline

1 Introduction

This paper contains an extract from the Version 5, January 16, 2002, of the working paper that describes the ongoing SEEDS (Strategic Evolution of Earth Science Data Systems) cost estimation model and coupled requirements / levels of service study. Its objective is to support the SEEDS Formulation team in estimating the life cycle costs of future ESE data service providers and supporting systems. This extract, an introduction followed by Section 6 from the working paper, focuses on the working list of data service provider types that (drawing upon the NewDISS Concept Paper as a starting point) has been developed as part of the cost estimation model effort. The data service provider types will provide a basis for 'binning' like data service providers in the cost model data base, allowing the cost estimate for a future data service provider to be based on similar existing cases.

Section 2 below presents the General Data Service Provider Model's breakdown of functional areas. Section 3 (reproducing Section 6 of the working paper) describes the ESE data service provider types, including a reference model functional area by functional area description of each one. The intent is to not only describe the types but highlight differences between them.

The term 'data service provider' is used herein as a broad, generic term for a site that performs all or a subset of the functions defined in the general data service provider reference model. Many well known actual data centers such as the Distributed Active Archive Centers or the NOAA data centers will perform a subset of the general list of functions, while some sites described as 'data service providers' for this study, e.g. MODAPS (as a sample of a science team processing facility that does not perform archive or general user distribution), are different in function from many well known data centers but fit within the framework of the data service provider reference model.

The general data service provider reference model will have subsets corresponding to the tentatively defined ESE data service provider types. (This approach has the advantage of allowing the future definition of additional data service provider types, or variations of the types defined herein, i.e. other possible subsets, within the same general framework.)

2 General Data Service Provider Reference Model - Functional Areas

This section describes the Data Service Provider Reference Model, a model of a data service provider in terms of a set of functional areas that, taken together, comprise the range of functions that a data center performs and the areas of cost that must be considered by the cost estimation model. The different types of data service providers would function within all or some of these areas. In each functional area, the cost estimation model will consider implementation and operations costs. Some of the areas are not strictly speaking "functional" in nature (such as 'facility / infrastructure') but are needed to ensure that all significant cost areas are included. The list of functional areas can be taken in toto as a summary of the areas of cost included in the model.

The following are working definitions of the functional areas that make up the data service provider reference model:

- **Ingest**—the process of receiving, reading, quality checking, cataloging, of incoming data to the point of insertion into the archive. Ingest can be manual or electronic with manual steps involved in quality checking, etc. Incoming data can be received from external sources or internally generated. Ingest can include format conversion, metadata extraction, or other preparation of incoming data for archive or use within the data service provider.
- **Processing**—the generation and quality checking of new derived data products from data or products that have been ingested, or previously generated, generally on a routine, operational basis. Processing includes process control (production planning, scheduling, monitoring, etc.) as well as product

generation per se. The data service provider may receive the software that embodies product generation algorithms from outside developers (e.g. some Terra instrument teams for the DAACs currently) who are responsible for the initial delivery and for delivering updated versions. Where quality, especially science quality, of products remains the responsibility of an outside developer, processing includes quality checking by the science software developer. Support provided by the data service provider for integration and test of this 'science software' is included as an activity under processing. In cases where a data service provider develops algorithm software, that effort (i.e. development, integration, and test) is included under Development.

- **Documentation**---the development (or upgrading of received) data and product documentation to meet adopted standards, including catalog information (metadata), user guides, etc., through consultation with data providers, algorithm developers, flight projects, etc.
- **Archive**—the insertion of data into archive storage, and management, handling and preservation of data, metadata, and documentation within a data service provider's archive. Inserted data can include data ingested from sources external to the site, or data/products generated on-site. Handling and preservation include quality screening of data entering and exiting the archive, quality screening of archive media, backups, and accomplishing migrations from one type of media to another. Insertion into the archive can be electronic or manual (e.g. hanging tapes on a rack or popping them into a robotic silo).
- **Distribution**—providing access to catalog information and a search and order capability to users, receiving user requests for data, fetching the requested data from the archive, performing any subsetting, reformatting / format conversion, or packaging, and providing the end product to the user by electronic means or on physical media. Catalog search and order can include providing local user interface and capability and/or providing an interface to a broader based, cross-site search and order capability (e.g. DAACs supporting search and order via the EOS Data Gateway).
- **User Support**—user support provided in direct contact with users by user support staff, including responding to queries, taking of orders, etc.
- **Instrument / Mission Operations** - monitoring instrument and spacecraft performance, generating instrument and spacecraft commands, and event scheduling (using NASA or other appropriate operational mission management services).
- **Sustaining Engineering** - Maintenance and enhancement of custom applications software (including any science software embodying processing algorithms developed by the site).
- **Engineering Support** - Some or all of the following as applicable at a particular site: systems engineering, test engineering, configuration management, coordination of hardware maintenance by vendors, COTS procurement, installation of COTS upgrades, system administration, database administration, network/communications engineering. Engineering support is internal, directed toward the internal operation of the data service provider.
- **Technical Coordination** - Includes participation in SEEDS system level processes, including coordination on standards and interfaces, work on common metrics, overall architecture, etc. Technical coordination, which by its nature includes engineering, is directed outward, supporting the data service provider as one element of a system of cooperating centers.
- **Implementation**---Includes development of, and making operational, the data and information system capabilities required by the data service provider to perform its mission, including design and implementation of the data system (hardware and system software) and applications software. Implementation can recur during the operating period as systems are expanded or replaced. In some cases applications software will include product generation software embodying science algorithms. Development can include development of software tools for use by users to unpack, subset, or otherwise manipulate products provided by the data service provider.
- **Management** - Includes management and administration at the data service provider level ("front office") and direct management of functional areas, and internal support: some or all of the following as applicable at a particular site: logistics, supplies, facilities, security management, property inventory and

management, facility management.

- **Facility / Infrastructure** - Includes a variety of non-staff cost factors such as supplies, facility lease and utility costs and other similar overhead costs, hardware maintenance, COTS licenses, etc.

3 ESE Data Service Provider Types

This section describes ESE data service provider types, drawing on the NewDISS concept paper "Draft Version 1.0 - NewDISS: A 6-to-10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise", October 2000 for a starting point. For each data service provider type, this section will present the conceptual description taken from the concept paper and a description of the functions of the data service provider type in terms of the data service provider reference model and its functional areas (which will define the subset of the reference model that applies to the data service provider type).

The NewDISS concept paper introduces its discussion of NewDISS data service provider types: "NASA's ESE has requirements for collection and synthesis of scientific information, for bringing synthesized data products to bear on unanswered scientific questions, and for preserving data and information for future scientific discovery. … NewDISS is therefore seen as consisting of a dynamic network of interconnected components, each responsive to its environment, containing capabilities for change over time through feedback with the science community. These components will be responsible for executing NewDISS data management functions and must allow easy participation by scientists and data and services providers. The components of NewDISS have been conceptualized (October, 2000) as including "Backbone" processing centers, PI-managed Mission Data Centers [here Mission Data Service Providers], Science Data Centers [here Science Data Service Providers], and Multi-Mission Data Centers."

Three additional data service provider types are added:

1. Applications Center, focused on uses and users other than research, given the existence of NASA funded applications activities such as Type III ESIPs and RESACS (see Section 7);

2. Information Center, focused on information describing data and products rather than the data and products themselves, based on discussion at the Formulation Team Retreat, November 7-8, 2001, where 'ECHO' was suggested as a possible future instance, and the GCMD is plainly a currently operational instance.

3. Long Term Archive Center, focused on permanent preservation and archiving of data and products and their documentation and active support to climate research, etc., based on a request from Matt Schwaller, a member of the Formulation Team and leader of the Earth Science Data Life-Cycle study. Long term archiving is strictly speaking not an ESE responsibility, but inclusion of a hypothetical Long Term Archive data service provider type is intended to support planning that NASA is doing with NOAA and USGS, the agencies who have (with NARA) the long term archive responsibility.

A "data service provider" does not necessarily imply a physically distinct institution. An institution such as a NASA center, a university, an organization of another US Government Agency such as USGS or NOAA can host a data service provider or a combination of data service providers. This is equivalent to the existing situation in which the University of Colorado hosts the NSIDC DAAC, or the USGS's EROS Data Center hosts the EDC DAAC.

3.1 Backbone Data Center

This section describes the generic Backbone Data Center.

3.1.1 Backbone Data Center Concept

The following is the concept for Backbone Data Centers, from the NewDISS Concept Paper: "These centers, most likely evolving from some of the current DAAC's, will address NASA's responsibility for preserving and protecting the large volumes of data from the ESE satellite missions. One of the primary roles of the backbone data centers will be to preserve the basic data. Clearly, NASA can provide a considerable amount of existing

infrastructure and technical skill needed to provide satellite mission data downlink and "level 0" or "level 1" data processing. Teaming NASA missions with Backbone Data Centers in the Announcement of Opportunity (AO) process for backup or for generation of basic data products may well be an attractive option for handling some of the core data management requirements of NewDISS. Another role for the Backbone Data Centers will be to acquire products agreed to be scientifically important for preservation and to prepare all these data for long-term archiving. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives. Backbone Data Centers, staffed by professional data managers, provide a core set of historical experience and proven capabilities. As such, they provide a means for risk mitigation against the failure of one or more of the NewDISS components by serving as backup centers for the other parts of the NewDISS. These data centers would most likely be few in number to ensure the cost-effectiveness of the NewDISS."

3.1.2 Backbone Data Center Functions

In general, the Backbone Data Center is expected to provide stable and highly robust services, with a key responsibility for data preservation and documentation, and with a mandate to provide professional data management as a resource for ESE as a whole. Backbone Data Centers are not identified with a particular mission or project but provide data management services in support of multiple missions and the NASA science program in general. Backbone Data Centers have an indefinite lifespan subject to regular review of their performance.

Backbone Data Centers could provide processing functions for NASA missions through teaming arrangements with NASA Principal Investigators, and can serve as a backup to other ESE data service providers.

The paragraphs below will discuss the Backbone Data Center role in each of the general data service provider reference model's functional areas.

Mission and Instrument Command and Control - The Backbone Data Center does not perform this function.

Ingest - The Backbone Data Center performs ingest of a wide variety of data types, ranging from low level data streams to ancillary data to all of the levels of derived products. In some cases the ingest function must be performed on a time critical, operational basis, e.g. for data and supporting information received from operating satellite platforms via NASA or other agency mission operations and communications systems. Quality control on incoming data is critical for lower level (e.g. level 0) data ingested, as the Backbone Data Center must detect bad data and request replacement data from operational sources that may have a limited capability for storing and retransmitting data.

Processing - The Backbone Data Center may perform processing through a teaming arrangement with a flight mission Principal Investigator, which can include large scale (in terms of number of products generated and /or product volume data) operational 'standard product' processing and reprocessing, perhaps with emphasis on Level 1 processing vs higher level derived product processing. Processing by the Backbone Data Center would be highly reliable with tight quality control.

Documentation - The Backbone Data Center ensures that its data and product holdings are documented to the standard for long term archiving, working as necessary with external data sources (e.g. other data service providers) to capture all needed information.

Archive - The Backbone Data Center provides a very robust archive capability, performing insertion of data into archive storage, and preservation of data, metadata, and documentation within the archive. Preservation measures should include quality screening of data entering and exiting the archive, quality screening of archive media, off-site backup with sampling to verify integrity, and accomplishing migrations from one type of media to another.

Distribution - The Backbone Data Center serves a broad user community with a robust search and order and distribution (electronic and media) service, including offering subsetting, reformatting, repackaging in response to user needs. The Backbone Data Center will also transfer data and documentation to designated long term archive centers in accordance with life cycle data management plans.

User Support - The Backbone Data Center provides effective user support for a wide range of users.

Sustaining Engineering - The Backbone Data Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Backbone Data Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - Internal support performed by the Backbone Data Center must include supporting an archive facility that is environmentally controlled and physically secure and a separate off-site backup archive.

Management - The Backbone Data Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.2 Mission Data Service Provider

This section describes the generic Mission Data Service Provider.

3.2.1 Mission Data Service Provider Concept

The following is the concept for NewDISS Mission Data Service Providers, from the NewDISS Concept Paper: "These data systems are specifically affiliated with instruments or satellite systems. They are either PI-led or facility/project-led. They provide key measurements and standard products from NASA-supported satellite instruments. The key characteristic of the mission data centers [here mission data service providers] is that they will be engineered and implemented as part of an ESE mission proposal. It is anticipated that these Mission Data Centers could leverage the activity at the current ESE data management infrastructure: the ECS flight operations and science data systems and the other hardware and software infrastructure at the DAAC's, the ESIP's, and the SCF's. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives. Mission Data Centers will also need to address satellite/instrument command and control and data downlink. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives, such as commercially provided or university support services.

Mission Data Service Providers will be responsible for their data management functions during an Earth-observation space flight mission. These data service providers will be funded by the mission selected through the ESE flight programs and will be selected by competitive selection for future ESE missions."

3.2.2 Mission Data Service Provider Functions

In general, the Mission Data Service Provider is an element of a particular ESE mission that exists to provide data management services for the life of that mission. The mission might be might involve an instrument on an independently operated spacecraft (such as SeaWinds on ADEOS) or might include multiple instruments on a dedicated spacecraft (such as Terra or Aqua). The services provided by the Mission Data Service Provider extend from instrument or platform command and control through generation and distribution to mission science team members of science products derived from instrument data for quality assurance, validation, and research. A Mission Data Service Provider would provide instrument data and science products to a Backbone Data Center for distribution to the broad user community and archive after the mission life is completed.

Mission and Instrument Command and Control - The Mission Data Service Provider performs this function for instruments and spacecraft that are part of its mission through NASA or other appropriate operational

mission management services. This includes monitoring instrument and spacecraft performance, generating instrument and (if applicable) spacecraft commands, and event scheduling.

Ingest - The Mission Data Service Provider ingests instrument and spacecraft telemetry and instrument data from NASA or other spacecraft operations and communications systems, and ancillary data needed to support product generation from various sources. Ingest of instrument data and instrument and spacecraft telemetry might be performed on a time critical, operational basis, and the Mission Data Service Provider must detect bad data and request replacement data from operational sources that may have a limited capability for storing and retransmitting data.

Processing - The Mission Data Service Providers will perform small to large scale (in terms of number of products generated and /or product volume data) 'standard product' processing and reprocessing. If the processing is performed to meet the needs of the mission science team only, it can be performed as the team requires. If the processing also must meet the needs of other missions (e.g. as ancillary products), science teams, or other users, it may be performed on an operational basis (especially once processing algorithms become stable). Processing by the Mission Data Service Provider would include tight quality control. The Mission Data Service Provider could team with a Backbone Data Center for processing service, especially if there is a requirement for routine, operational generation of standard products.

Documentation - The Mission Data Service Provider generates complete documentation of its instrument data and all derived products. The Mission Data Service Provider cooperates with a Backbone Data Center that receives its data after completion of its mission to ensure that documentation is brought to long term archiving standards.

Archive - The Mission Data Service Provider would not perform an archive function per se, but would maintain secure working storage of data and products until their transfer to a Backbone Data Center at some time during the mission or after completion of the mission. The Mission Data Service Provider would maintain an off-site back up of all data for which it is responsible, and might use the services of a Backbone Data Center for this purpose.

Distribution - The Mission Data Service Provider provides products to the mission science team for quality assurance, validation, or research, with a search and order capability as needed to meet the needs of the mission science team. The Mission Data Service Provider will also transfer data, products, and documentation to a Backbone Data Center either during its mission as backup or when broader distribution of its data and products is appropriate, or at the conclusion of the mission.

User Support - The Mission Data Service Provider provides close support to member of the mission science team.

Sustaining Engineering - The Mission Data Service Provider performs sustaining engineering, with no or very infrequent interruption of any critical operational capabilities.

Engineering Support - The Mission Data Service Provider performs engineering support functions as needed, but with no or very infrequent interruption of any critical operational capability.

Internal Support - Internal support performed by the Mission Data Service Provider must include supporting a working storage facility that is environmentally controlled and physically secure and a separate off-site backup, for which the Mission Data Service Provider might use the services of a Backbone Data Center.

Management - The Mission Data Service Provider provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.3 Science Data Service Provider

This section describes the generic Science Data Service Provider.

3.3.1 Science Data Service Provider Concept

The following is the concept for NewDISS Science Data Service Provider, from the NewDISS Concept Paper: "These data centers [here data service providers] will collect data from multiple missions for a user community focused on a single research question. There are several examples of these types of Science Data Centers in NASA's Space Science Enterprise. These centers are targeted at specific science questions (perhaps from the NRC Pathways Report) and/or science disciplines, and they directly support research and data analysis for specific research questions. These data centers will address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives."

3.3.2 Science Data Service Provider Functions

In general, the Science Data Service Provider is a temporary data management capability implemented to support a particular research effort by a limited community of users (which will be called its 'research team'). The Science Data Service Provider operates in a research environment, without the need for robustness and performance as would be the case for an operational environment.

Mission and Instrument Command and Control - None.

Ingest - The Science Data Service Provider obtains data and products required to meet the research objectives of its research team from a variety of sources, including other ESE data service providers, other agency data centers, etc. The ingest would not be performed on a time critical, operational basis.

Processing - The Science Data Service Provider would perform processing, and in some cases reprocessing, of new science products developed by the research team on an ad hoc basis.

Documentation - The Science Data Service Provider generates complete documentation of its science products. The Science Data Service Provider cooperates with a Backbone Data Center that receives its products after completion of its working life (or with the designated long term archive for its products) to ensure that documentation is brought to long term archiving standards.

Archive - The Science Data Service Provider would not perform an archive function per se, but would maintain working storage of products obtained from other sources or science products generated as part of the research effort it supports.

Distribution - The Science Data Service Provider generates complete documentation any new science products developed by the research team that constitute new research quality products to be made available to the general science community (e.g. products cited in publications by members of the research team which should be available other scientists seeking to corroborate or extend the research performed by the team). The Science Data Service Provider will make the products collected to support the research effort readily available to members of the research team, and will perform reformatting, subsetting, or packaging of those products as needed to facilitate their interuse by the research team. The Science Data Service Provider will also transfer new research quality science products and documentation to a Backbone Data Center when broader distribution of those products is appropriate, or at the conclusion of the research effort.

User Support - The Science Data Service Provider provides close support to member of the research team it supports.

Sustaining Engineering - The Science Data Service Provider performs software maintenance as needed.

Engineering Support - The Backbone Data Center performs engineering support functions as needed.

Internal Support - Internal support performed by the Science Data Service Provider includes supporting a working storage facility and a separate off-site backup of any new research quality science products generated by the research effort (e.g. that are cited by research team publications), for which the Science Data Service Provider might use the services of a Backbone Data Center.

Management - The Science Data Service Provider provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.4 Multi-Mission Data Centers

This section describes the generic Multi-Mission Data Center.

3.4.1 Multi-Mission Data Center Concept

The following is the concept for NewDISS Multi-Mission Data Centers, from the NewDISS Concept Paper: "A fourth type of data center [here data service provider] is the Multi-Mission Data Center. An example of the type of data activity to be carried out by such a data center is the generation of consistent time-series geophysical parameters, an activity exemplified by the current National Oceanic and Atmospheric Administration (NOAA)/NASA Pathfinder Datasets program, which is funded by NASA's ESE and carried out by PIs at various institutions. These efforts will take on more importance in the future, since NASA ESE has the requirement for generating time-series of geophysical parameters, while the EOS mission strategy has evolved so that it is now designed to accommodate technological change. Thus, these efforts will include construction of the long-time scale datasets from more than one NASA (or other) mission. These data centers will need to address network connectivity as part of their on-going activities. Selection of these services will be driven by PI-teaming arrangements, using either NASA-available resources or competitive alternatives."

3.4.2 Multi-Mission Data Center Functions

In general, the Multi-Mission Data Center is a temporary but potentially long lived data management capability implemented to support a particular data synthesis effort by a limited community of users (which will be called its 'synthesis team'). An example of a data synthesis effort would be research into how to cross-calibrate and consistently map measurements made by different missions (perhaps overlapping or consecutive) in order to be able to generate a consistent, continuous, long-term, research quality data set spanning multiple instruments/missions, validation of the cross-calibrated data sets, and then the production of the long time series data set. Such a production effort could be quite intensive in order to accomplish in a reasonable time the generation of a long time series data set involve handling many year's worth of a number of good sized data sets. The Multi-Mission Data Center operates in a research environment, without the need for robustness and performance as would be the case for an operational environment.

The distinction drawn between a Science Data Service Provider and a Multi-Mission Data Center is that the former supports a particular research effort, while the latter supports a data synthesis effort that would enable future science efforts using the new, research quality data sets it produces.

Mission and Instrument Command and Control - None.

Ingest - The Multi-Mission Data Center obtains data and products required to meet the research objectives of its synthesis team from a variety of sources, including other ESE data service providers, other agency data centers, etc. The ingest would not be performed on a time critical, operational basis, but could involve large amounts of data if long time series of large data sets are involved.

Processing - The Multi-Mission Data Center would perform processing of new data synthesis products (such as long time series data sets) developed by the synthesis team on an ad hoc basis. This processing could be a major effort, for example if the objective is a long time series product produced from a number of large, multi-year input data sets. The Multi-Mission Data Center could accomplish a large scale processing effort (such as a major effort to generate a long time series data set once the cross-calibration, mapping, etc., involved had been tested and validated) through a partnership with a Backbone Data Center or other processing facility.

Archive - The Multi-Mission Data Center would not perform an archive function per se, but would maintain working storage of data and products obtained from other sources and new data synthesis products generated by the center. This could involve large data volumes, and the working storage would be configured to facilitate the processing effort.

Distribution - The Multi-Mission Data Center generates complete documentation any new data synthesis products developed by the synthesis team that are new research quality products to be made available to the general science community, including full, documentation of the cross-calibration and any other steps taken to build the consistent time series. The Multi-Mission Data Center will make the products collected to support the data synthesis effort readily available to members of the synthesis team. The Multi-Mission Data Center will also transfer new research quality data synthesis products and documentation to a Backbone Data Center when broader distribution of those products is appropriate, or at the conclusion of the data synthesis effort.

User Support - The Multi-Mission Data Center provides close support to member of the synthesis team it supports.

Sustaining Engineering - The Multi-Mission Data Center performs software maintenance as needed.

Engineering Support - The Backbone Data Center performs engineering support functions as needed.

Internal Support - Internal support performed by the Multi-Mission Data Center includes supporting a working storage facility and a separate off-site backup of any new research quality data synthesis products generated by the synthesis effort, for which the Multi-Mission Data Center might use the services of a Backbone Data Center.

Management - The Multi-Mission Data Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.5 Applications Center

This section describes the generic Applications Center.

3.5.1 Applications Center Concept

NewDISS Applications Centers will obtain NASA Earth science products and use these, sometimes in conjunction with other Earth science data or any kind of other data to produce special products and/or deliver tailored services to an applications community. These communities could include agriculture, fisheries, urban planning, resource management, many etc., which could derive value from NASA Earth science products if they were suitably formatted or packaged or used in conjunction with other data to produce new products specifically designed to meet the needs of the application community. Examples of existing ESE applications activities include the Socio-Economic Applications Data Center (SEDAC), the Type-III ESIPs, and RESACs, all of which are discussed in Section 7 below.

3.5.2 Applications Center Functions

In general Applications Centers perform the same functions as other data service provider types, the primary distinction being the nature of their user community and therefore their products and services. Thus the Applications Center type actually embraces a wide variety of possible functional models, ranging from large operational activities to small activities providing very focused support to a small user community. Consequently cost modeling of Applications Centers would be as complex as the modeling of all of the other types of ESE data service providers - or, in other words, one could identify a range of sub-types within the spectrum of Applications Centers which would be distinct subsets of a generalized applications center type. This level of complication could be pursued in the future. The best approach to take can be considered once a better understanding of the existing array of applications centers (see Section 7 as it develops) is reached.

Applications Centers can be ESE funded for their lifetime, can be given temporary start-up funding to support their development into self sustaining entities, or can be self funding entities that partner with ESE.

Mission and Instrument Command and Control - None.

Ingest - The Applications Center obtains data and products required as inputs for its applications products from other ESE data service providers, other agency data centers, etc. In some cases the ingest would be

performed on a time critical, operational basis, and in other cases might be on an ad hoc or intermittent basis, and could involve large amounts of data.

Processing - The Applications Center would perform processing of new applications products (such as products for agriculture or fisheries) developed by the Applications Center. This processing could be a major effort if low level data sets of large size are used to generate products on a routine basis.

Documentation - The Applications Center would generate documentation sufficient to support the current use of its products.

Archive - The Applications Center would not be likely to perform an archive function per se, depending perhaps on the commercial value of its products beyond their first use, but would maintain working storage of data and products obtained from other sources and new applications products generated by the center. This could involve large data volumes, and the working storage would be configured to facilitate the processing effort.

Distribution - The Applications Center may distribute its products to either a very limited user community or a very broad user community, operationally or intermittently or on an ad hoc basis depending on its particular mission or business plan.

User Support - The Applications Center provides support to its users according to its particular needs, which will vary considerably from case to case.

Sustaining Engineering - The Applications Center performs software maintenance as needed.

Engineering Support - The Applications Center performs engineering support functions as needed.

Internal Support - The Applications Center performs internal support functions as needed.

Management - The Applications Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.6 Information Center

This section describes the generic ESE Information Center.

3.6.1 Information Center Concept

In general the Information Center performs many of the same functions as the Back Bone Data Center, except that the Information Center is concerned with information describing data and products (i.e., one or more types of metadata) rather than the data and products themselves. In general the Information Center will obtain its information from other data service providers, assemble it and make it available to its users, and when its users discover data or products they desire, then help (e.g. by providing links to data service provider websites) those users obtain access to the services of source data service providers.

The addition of this data service provider type was based on discussion at the Formulation Team Retreat, November 7-8, 2001, where 'ECHO' was suggested as a possible future instance. The GCMD is a currently operational instance.

3.6.2 Information Center Functions

The paragraphs below will discuss the Information Center role in each of the general data service provider reference model's functional areas.

Mission and Instrument Command and Control - The Information Center does not perform this function.

Ingest - The Information Center performs ingest of one or more metadata types, ranging from product instance (e.g. granule) level inventory metadata streams to overall product type descriptions or service descriptions. In some cases the ingest function may be performed on a time critical, operational basis, e.g. for inventory metadata received from other data service providers to be posted to the Information Center's inventory on an

operational basis. In other cases, ingest of product type descriptions (etc.) are received on an ad hoc basis and are infrequently updated. Quality control on incoming metadata is critical if the Information Center's database is to be current with consistent and accurate content.

Processing - The Information Center does not perform this function.

Documentation - The Information Center ensures that is own content is consistent and complete but does not generate or maintain any other documentation..

Archive - The Information Center provides only working storage for its database of descriptive information.

Distribution - The Information Center serves a broad user community with a robust metadata search capability. While its 'distribution' is of its own metadata, the Information Center facilitates access to the data and products its metadata describes. This might be in the form of links to source data service provider websites, or the ability to accept a user request for relay to a source data service provider.

User Support - The Information Center provides effective user support for a wide range of users who access its metadata holdings and to the source data service providers who provide the metadata.

Sustaining Engineering - The Information Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Information Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - The Information Center performs internal support functions with no or very infrequent interruption of its operations.

Management - The Information Center provides management for its own operation and staff to support its participation in SEEDS system level activities.

3.7 Long Term Archive Center

This section describes the generic Long Term Archive Center.

3.7.1 Long Term Archive Center Concept

The report, "Global Change Science Requirements for Long-Term Archiving" (USGCRP, March 1999), of the results of the science panel that met in a workshop held at NCAR in October, 1998, discussed the essential functions and characteristics of a long term archiving program.

In general the Long Term Archive Center performs most if not all of the same functions as the Backbone Data Center, with the additional focus on permanent preservation and archiving of data and products and their documentation, and active support to climate research, etc., that requires reprocessing of and/or access to long time series of data and products. The Long Term Archive Center participates with ESE data service providers in life cycle data management planning and in a process for obtaining science guidance and priorities for long term archiving.

Long term archiving is strictly speaking not an ESE responsibility, but inclusion of a hypothetical Long Term Archive Center type is intended to support planning that NASA is doing with NOAA and USGS, the agencies who have (with NARA) the long term archive responsibility.

3.7.2 Long Term Archive Center Functions

The paragraphs below will discuss the Long Term Archive Center role in each of the general data service provider reference model's functional areas, drawing on the USGCRP report cited in Section 3.7.1 above. Items in the functional discussion below that are explicitly derived from that report are indicated by an appended '(USGCRP)'.

Mission and Instrument Command and Control - The Long Term Archive Center does not perform this function.

Ingest - The Long Term Archive Center performs ingest of a wide variety of data and product types, ranging from low level data streams to ancillary data to all of the levels of derived products, and their documentation. These products may be new to the center or may be replacements of earlier versions of products already archived by the center.

It is essential that the Long Term Archive Center verify the integrity and quality of data and derived product and associated documentation as it is ingested into the archive (USGCRP).

The ingest would be a transfer from another data service provider, e.g. a Backbone Data Center, according to scenario to be documented in life cycle data management plans. If the transfer is from a research environment (e.g. a Science Data Service Provider) that Long Term Archive Center should proactively reach out to the research source and develop the needed agreements and procedure, assist in planning documentation, etc., (USGCRP). The transfer could be a single bulk delivery, or staged as a series of deliveries over a period of time. The transfer could be by media or network.

Processing - It is essential that the Long Term Archive Center exercise data to produce new products and/or new versions of old products to validate data and product documentation, identify and resolve problems in the data, provide opportunities to scientists within the center to pursue science interests, produce new or updated products that are of value to the science community, provide an opportunity to rethink and reorganize how the data are stored to take into account user access needs as well as accommodate new storage and access technology, and increase data longevity (USGCRP). Typical science processing / reprocessing efforts could include production of long time series of intercalibrated data sets from multiple sources/ sensors to support climate change research.

Processing / reprocessing by the Long Term Archive Center would be on an ad hoc basis, but with tight quality control.

Documentation - It is most essential that the Long Term Archive Center ensure that its data sets and products in the archive are accompanied by complete, comprehensive, and accurate documentation (USGCRP), in accordance with long term archive documentation standard. The center works as necessary with external data sources (e.g. other data service providers) to capture all needed information.

Archive - The Long Term Archive Center provides a very robust archive capability, performing insertion of data into archive storage, and preservation of data, metadata, and documentation within the archive. Preservation and maintenance of data holdings, including ensuring integrity and quality of the data, products, and associated documentation is an essential function of the Long Term Archive Center (USGCRP). Extension of maintenance to include updating of documentation with user comments on the data or product is desirable (USGCRP).

Preservation measures should include quality screening of data entering and exiting the archive, quality screening of archive media, off-site backup with sampling to verify integrity, and accomplishing migrations from one type of media to another. It is essential that the Long Term Archive Center develop and maintain a multi-year data migration plan, and that the center perform integrity checks on archive media between migrations (USGCRP).

Data migrations to new archive technology should be taken as opportunities for processing / reprocessing (USGCRP).

Distribution - The Long Term Archive Center serves a broad user community with a robust search and order and distribution (electronic and media) service, including offering subsetting, reformatting, repackaging in response to user needs. It is essential that the center provide the next and subsequent generation of scientists with appropriate access to, and facilitate their use of, its holdings, where 'access' includes a data set / product search and order function, the ability to deliver data and/or products and supporting information

(documentation) on suitable media or electronically, and choices of format, user options such as subsetting, that facilitate access and use (USGCRP).

User Support - The Long Term Archive Center provides effective user support (a user support staff knowledgeable about the data and products, willing and able to help users identify, obtain, and use the products the need, including making referrals to other sources of data - USGCRP) for a wide range of users.

Sustaining Engineering - The Long Term Archive Center performs sustaining engineering, with no or very infrequent interruption of operational capabilities.

Engineering Support - The Long Term Archive Center performs engineering support functions with no or very infrequent interruption of its operations.

Internal Support - Internal support performed by the Long Term Archive Center must include supporting an archive facility that is environmentally controlled and physically secure and a separate off-site backup archive.

Management - The Long Term Archive Center provides management for its own operation and staff to support its participation in archive related activities. For example, it is essential that the center be actively facilitate the process for deciding which products to include or exclude from, or remove from, the archive (USGCRP). It is essential this process be driven by science priorities and scientific assessments, and that scientists be actively engaged in the process: setting criteria and making decisions (USGCRP). The Long Term Archive Center would participate with the appropriate ESE data service providers in these processes.

References

1. "NewDISS: A 6-to-10-year Approach to Data Systems and Services for NASA's Earth Science Enterprise - Draft Version 1.0", October 2000.

2. "NewDISS Level 0 Requirements", September 2001, Vanessa Griffin ESDIS/SOO and SEEDS Formulation Team.

3. "ESDIS Project Level 2 Requirements: Volume 5: EOSDIS Version 0, Revision B", March 2000, GSFC.

4. "ESDIS Data Center Best Practices and Benchmark Report", September 2001, SGT Inc.

5. "Ensuring the Climate Record from the NPP and NPOESS Meteorological Satellites", NRC Committee on Earth Studies, September 2000.

6. "Global Change Science Requirements for Long-Term Archiving", NOAA-NASA and USGCRP Program Office, March 1999.

Others TBD.